

DOI:10.12158/j.2096-3203.2022.01.012

# 基于故障链聚类算法的电网关键线路辨识

黎寿涛<sup>1,2</sup>, 夏成军<sup>1,2</sup>, 钟明明<sup>1,2</sup>, 管霖<sup>1,2</sup>

(1. 华南理工大学电力学院, 广东 广州 510640; 2. 广东省新能源电力系统智能运行与控制企业重点实验室, 广东 广州 510663)

**摘要:**研究连锁故障发展机理以及辨识故障演化路径中的关键线路,对揭示电力系统薄弱环节、降低连锁故障风险具有重要意义。为此,文中提出一种基于故障链聚类算法的电网关键线路辨识方法。首先,建立改进的基于直流潮流法的电力系统解列模拟器(DCSS)连锁故障仿真模型,配合随机化学(RC)法高效生成含详细时序信息的故障链集合;然后,以编辑距离为相似性指标实现对故障链的层次聚类,分类后的故障链集合不仅可以降低后续数据挖掘的难度,而且能够更加全面地识别出不同连锁故障演化模式中的薄弱环节;最后,以 Matpower 2 383 节点系统为例,根据关键线路扩容后系统风险水平的下降程度评估各类算法所辨识线路的重要性。结果表明文中方法能更好降低系统连锁故障风险水平,证明了所提模型及算法的有效性。

**关键词:**连锁故障;相似性度量;编辑距离;层次聚类;风险评估;关键线路辨识

中图分类号:TM711

文献标志码:A

文章编号:2096-3203(2022)01-0084-09

## 0 引言

近年来,国内外发生了多起由连锁故障引发的大停电事故<sup>[1-3]</sup>,造成了巨大的经济损失与恶劣的社会影响。研究表明,在连锁故障的发展过程中极少数线路有至关重要的作用<sup>[4]</sup>,高效且准确地识别出这些关键线路,对于揭示电网薄弱环节、提高电力系统可靠性具有重要的意义。

基于复杂网络理论的关键线路辨识算法首先将电力系统简化为抽象模型,再依据节点介数<sup>[5]</sup>、线路效能权值<sup>[6]</sup>、潮流熵<sup>[7]</sup>等拓扑特征对线路进行重要性排序。此外,亦有学者建立了如 ORNL-PSERC-Alaska(OPA)模型<sup>[8]</sup>、Manchester 模型<sup>[9]</sup>、基于直流潮流法的电力系统解列模拟器(direct current power flow simulator of power system separation, DCSS)模型<sup>[10]</sup>等连锁故障仿真模型,试图刻画电力系统连锁故障的演化过程,并从中找出对故障发展具有推波助澜作用的关键线路。

上述仿真模型或电力系统实际运行所产生的故障链数据蕴含了丰富的连锁故障信息,通过分析不同故障链间的共同特征及发展规律,同样能实现关键线路的辨识。文献[11]利用 Apriori 算法对交直流连锁故障链进行频繁项挖掘,有效地识别出危及系统安全的强关联规则;文献[12]通过 FP-growth 算法对故障链集合进行数据挖掘,辨识出了不同故障发展过程中的危险线路;文献[13]根据故障链集

合构建了连锁故障时空图,并从触发与扩大连锁故障 2 个角度出发,分类辨识出了系统中的脆弱线路。然而,故障链集合中的数据可能由不同的连锁故障演化模式产生,各模式下起主导作用的关键线路亦不尽相同<sup>[13]</sup>。若在进行关键线路辨识前,预先对故障链序列的相似性进行度量,根据序列间相似的演化规律或关联特征进行聚类,并对相似度较高的故障链集合进行分类评估,可进一步提高关键线路辨识精度与效率<sup>[14]</sup>。

文中首先根据热量累积效应构建了改进 DCSS 连锁故障仿真模型,并与随机化学(random chemistry, RC)法相互配合,高效生成了含丰富时序信息的故障链集合。在此基础上,文中引入编辑距离(edit distance, ED)衡量故障链间的相似程度,并通过凝聚式层次聚类算法实现了故障链集合聚类。将完成聚类后的故障链集合分别进行线路风险重要度排序,名次靠前的线路即为相应连锁故障演化模式中起主导作用的关键线路。以 Matpower 2 383 节点系统为例,对不同算法所辨识的关键线路进行容量扩建,并根据扩容后系统连锁故障风险水平的下降量化比较了所辨识线路的重要程度,对比结果进一步证明了文中所提模型及算法的有效性。

## 1 连锁故障仿真模型

### 1.1 故障链集合生成

DCSS<sup>[10]</sup>是一种基于直流潮流法建立的连锁故障仿真模型,主要面向过载主导型连锁故障,考虑在无功率较为充足的纯交流系统中因线路过载而发

收稿日期:2021-09-08;修回日期:2021-11-28

基金项目:国家自然科学基金资助项目(52077080)

生相继失效的连锁故障过程,其仿真速度快且没有收敛性问题。DCSS 模型通过对线路过载功率进行时间积分,模拟反时限保护机制,但当重载线路处于临界状态时,累计积分值可能被多次清零,所求线路跳闸的用时变长,使得模型所描述的故障时序特性可能与实际情况存在较大出入。因此,文中根据热量累积效应<sup>[15]</sup>建立更合理的改进 DCSS 连锁故障仿真模型。

首先,假设线路  $L$  的潮流  $P_L$  超过限值  $P_{\max,L}$ , 根据线路流过潮流对导线的热效应,可近似计算出线路  $L$  在  $t$  时刻的瞬时温度  $T_L(t)$ 。

$$\begin{cases} T_L(t) = e^{-\nu t} (T_L(0) - T_e(P_L)) + T_e(P_L) \\ T_e(P_L) = \frac{\alpha}{\nu} P_L^2 + T_{\text{env}} \end{cases} \quad (1)$$

式中:  $T_L(t)$  为线路的瞬时温度;  $P_L$  为线路  $L$  流过的功率;  $T_e(P_L)$  为线路流过该功率时的平衡温度;  $\nu, \alpha$  为线路热效应的相关参数,皆取决于线路本身;  $T_{\text{env}}$  为环境温度;  $T_L(0)$  为初始温度。

将线路过热保护的溫度触发閾值  $T_{\max,L}$  代入式(1),可算得线路  $L$  的过热保护动作时间  $\Delta t_{\text{trip},L}$ 。

$$\Delta t_{\text{trip},L} = -\frac{1}{\nu} \ln \left( \frac{T_{\max,L} - T_e(P_L)}{T_L(0) - T_e(P_L)} \right) \quad (2)$$

在给定电网参数的前提下,通过式(2)可快速计算出下一次故障跳闸线路及相应动作时刻。

上述所建立的改进 DCSS 模型能高效且准确地生成故障链集合,具体流程见图1,具体步骤为:

(1) 载入电网模型,初始化电网的运行状态。

(2) 设置初始故障,将相应故障线路作为初始扰动进行开断,第  $j$  条故障链  $C_j$  的初始故障线路集合为  $L_{\text{init},j}$ 。

(3) 通过孤岛判定程序对线路开断后的电网进行孤岛判定。假如有新的孤岛生成则进行发电机有功出力调整,如不能在规定时间内实现负荷与出力之间的平衡,则进行切机切负荷操作;假如没有新的孤岛出现,则直接转至步骤(4)。

(4) 对每一个仍处于运行状态的孤岛执行直流潮流运算。

(5) 依据步骤(4)所算得的潮流结果判断是否有线路过载,如果仍有过载线路,则转至步骤(6),否则转至步骤(7)。

(6) 将过载线路的潮流  $P_L$  代入式(2),计算下一次线路动作时间  $\Delta t_{\text{trip},L}$ ,相应开断线路记为  $L_{\text{next}}$ ,总仿真时间跳转至  $t_{\text{next}}$ 。记录跳闸线路与相应的动作时刻  $\{L_{\text{next}}, t_{\text{next}}\}$  后跳转至步骤(3)。

(7) 连锁故障仿真过程结束,根据步骤(6)所

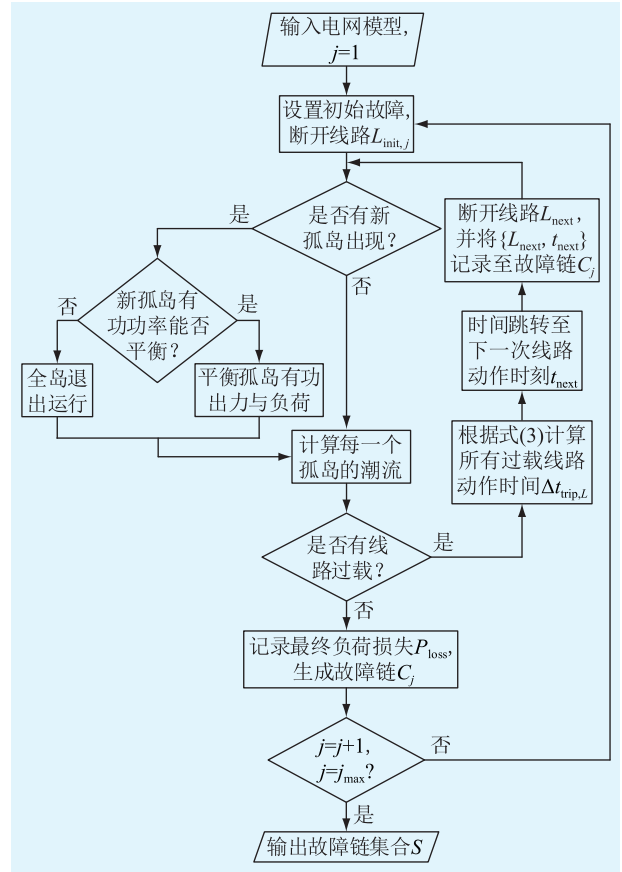


图1 故障链集合生成流程

Fig.1 Flow chart of fault chain set generation

记录的所有线路开断信息  $\{L_{\text{next}}, t_{\text{next}}\}$  及最终负荷损失  $P_{\text{loss}}$  生成故障链  $C_j$ 。

(8) 重新选取初始故障线路形成初始故障集  $L_{\text{init},j}$ ,并重复步骤(2)~(7)生成故障链  $C_j$ ,直至故障链样本足够多或满足仿真退出条件。

对于较小的系统,或对计算速度无较高要求的搜索过程,初始故障集  $L_{\text{init},j}$  可设定为系统中所有线路组合,通过遍历法来进行连锁故障分析,但当电网规模较大时,组合规模及计算量急剧增加,将面临“维数灾”的挑战。为兼顾搜索时间与搜索精度,文中采用效率较高的 RC 法<sup>[16]</sup>来提高故障链的生成效率。RC 法属于无偏抽样,即所得的初始故障集不依赖于任何先验的概率或者指标,生成的故障链集合样本较为全面地涵盖了多种故障演化路径,为后续的聚类分析及数据挖掘提供了更加丰富的数据样本。

如图2所示,某些连锁故障仿真模型按故障演化阶段进行线路开断。以 OPA 快动态过程模型<sup>[8]</sup>生成的故障链如图2(a)所示,其同一阶段中多条线路开断,不区分先后顺序,此类故障链易造成无效规则的误筛,增大后续数据挖掘难度。相比之下,如图2(b)所示的通过改进 DCSS 模型仿真生成的

故障链物理含义明确,包含更丰富的时序信息,既有线路开断的先后顺序,又记录了相应的开断时刻,能较真实地刻画连锁故障的演化过程,为后续分析提供更合理的样本依据。

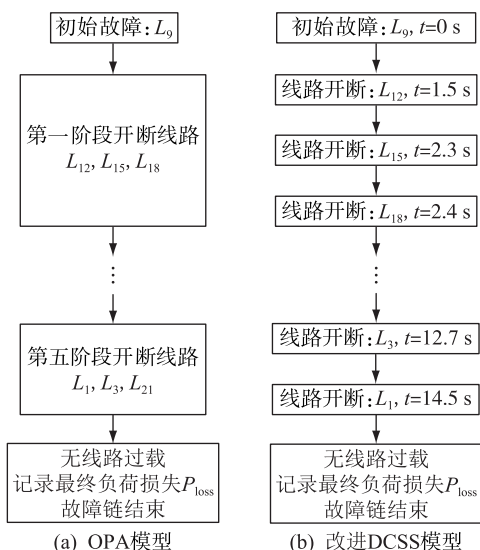


图2 2类连锁故障仿真模型所生成故障链的对比  
Fig.2 Comparison of fault chains generated by two kinds of cascading failure models

### 1.2 故障链的风险评估

文中采用文献[17]中的概率风险法进行故障链的风险评估,定义故障链  $C_j$  造成的后果与发生概率分别为  $E(C_j), p(C_j)$ 。

#### 1.2.1 连锁故障后果

连锁故障后果通常以系统负荷总损失、电网解列程度、故障链长度(停电线路数)等指标来衡量。负荷损失是连锁故障在电力系统中造成的最直接后果,因此文中采用系统负荷总损失来衡量连锁故障的后果。

$$E(C_j) = P_{\text{loss},j} \quad (3)$$

式中:  $P_{\text{loss},j}$  为第  $j$  条故障链  $C_j$  最终所造成的系统负荷总损失。

#### 1.2.2 连锁故障概率

根据 1.1 节中的定义可知,改进 DCSS 模型对于任意给定的  $N-k$  初始故障能且仅能生成唯一的故障链  $C_j$  及相应的负荷损失  $P_{\text{loss},j}$ ,即连锁故障的发生概率由  $N-k$  初始故障所决定,文中定义故障链概率  $p(C_j)$  为:

$$p(C_j) = \prod_{i=1}^k U(L_{\text{init},i}) \quad (4)$$

式中:  $k$  为初始故障元件数;  $L_{\text{init},i}$  为初始故障集  $L_{\text{init},j}$  中的第  $i$  线路;  $U(L_{\text{init},i})$  为线路  $L_{\text{init},i}$  的不可用率,可通过式(5)求得。

$$U(L_{\text{init},i}) = \lambda t_{\text{Dur}} / 8\ 760 \quad (5)$$

式中:  $\lambda$  为线路的平均失效频率;  $t_{\text{Dur}}$  为平均修复时间。独立事件的不可用率乘积表示多条线路同时失效的概率,即为以  $N-k$  故障为初始故障的连锁故障发生概率。

将上述的故障链后果和故障链概率相乘,则可计算出事故链的负荷损失期望值  $R(C_j)$ ,用于表征事故链的风险。

$$R(C_j) = E(C_j)p(C_j) \quad (6)$$

故障链集合  $S$  的连锁故障风险等于所含全部故障链的风险之和。

$$R(S) = \sum_{C_j \in S} R(C_j) \quad (7)$$

## 2 基于 ED 的故障链层次聚类

### 2.1 ED 的定义

在时间序列数据挖掘的诸多任务和问题中,序列相似性度量是最基础的问题。以动态时间归整(dynamic time warping, DTW)算法为代表的时序列度量方法已广泛应用在电力系统中,其通过对电流、电压等时间序列的波形轮廓进行匹配,完成故障源辨识<sup>[18]</sup>、扰动分类<sup>[19]</sup>等任务。而改进 DCSS 模型所生成的故障链本质是事件序列,如图 2(b)所示,此类序列具有不等长、离散性等特点,因此文中引入更直观的 ED 算法,计算故障链事件序列间的相似程度,为后续的分类过程提供聚类依据。

ED<sup>[20]</sup>是指将某一序列  $C_x$  通过插入、删除与替换 3 种编辑操作转换至另一序列  $C_y$  所需要的最小操作步数。假设存在 2 条由改进 DCSS 模型生成的故障链  $C_x, C_y$ ,按线路先后开断的次序可表示为:

$$\begin{cases} C_x = \{L_{x,1}, L_{x,2}, \dots, L_{x,m}\} \\ C_y = \{L_{y,1}, L_{y,2}, \dots, L_{y,n}\} \end{cases} \quad (8)$$

式中:  $L_{x,1}$  为故障链序列  $C_x$  中第 1 条开断的线路;  $m$  为  $C_x$  中开断线路的总数( $C_y$ 同理)。  $C_x$  与  $C_y$  之间的 ED 可通过构造一个  $(m+1) \times (n+1)$  维矩阵  $D$  递归求取,矩阵元素  $d_{i,j}$  的计算公式为:

$$d_{i,j} = \min(d_{i-1,j} + 1, d_{i,j-1} + 1, d_{i-1,j-1} + L_{x,i} \oplus L_{y,j}) \quad (9)$$

特别地,当  $i=0$  或  $j=0$  时,  $d_{i,j} = \max(i, j)$ 。其中,  $\oplus$  表示异或运算,相等时取 0,不等时取 1。最终算得矩阵  $D$  的右下角元素  $d_{m,n}$  即为故障链序列  $C_x$  与  $C_y$  之间的 ED,  $d_{m,n}$  越大代表  $C_x$  与  $C_y$  之的差异性越大。

当故障链序列长度不一致的时候,长序列之间的 ED 将远大于短序列之间的 ED,因此需要对 ED 进行归一化处理:

$$d'_{m,n} = \frac{d_{m,n}}{\min(m,n)} \quad (10)$$

归一化后的 ED 能较为准确地度量具有不等长、离散性等特点的故障链相似性,提高后续聚类分析的准确性与数据挖掘的效率。

## 2.2 基于 AGNES 算法的故障链聚类

文中采用凝聚式层次聚类 (AGglomerative NESTing, AGNES) 算法<sup>[21]</sup>,根据 2.1 节所得的故障链间的 ED,完成对故障链集合的聚类分析。

AGNES 算法“自底而上”的聚类过程如图 3 所示。随着 ED 的增加,故障链序列逐渐合并,直至最后聚至同一类中。聚类簇数可通过改变截取阈值来灵活调整,而最优聚类簇数可通过 Calinski-Harabaz (CH) 指标<sup>[22]</sup>进行评估求取。其中,  $S$  为由故障链  $C_j$  组成的故障链集合。

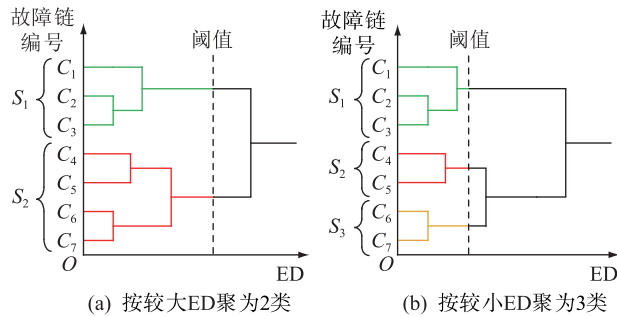


图 3 层次聚类

Fig.3 Tree diagram of hierarchical clustering

以 CH 指标较高的簇数对故障链集合进行划分后,同集合内故障链相似度较高,即线路开断的先后次序相近,则可近似认为此类故障链在同一连锁故障演化模式下产生。对各类故障链集合分别进行关键线路辨识,找出各连锁故障演化模式下具有推波助澜作用的薄弱环节,能有效提高辨识精度。

## 3 连锁故障关键线路辨识

### 3.1 线路风险重要度

假设故障链集合  $S$  共有  $m$  条故障链,其中有  $h$  ( $h \leq m$ ) 条故障链包含线路  $L_i$ ,即满足:

$$L_i \in C_j \quad j = 1, 2, \dots, h \quad (11)$$

则线路  $L_i$  的风险重要度的计算公式为:

$$I(L_i) = \sum_{j=1}^h (\omega(L_i, C_j) E(C_j)) \quad (12)$$

式中:  $\omega(L_i, C_j)$  为线路  $L_i$  在故障链  $C_j$  中的重要度权重。考虑到最终负荷损失是由故障链中所有线路开断造成的,且连锁故障演变过程存在级联性,因此认为先开断线路对负荷损失的影响高于后开断线路,相应权重也更高。文中假设重要度权重随开

断顺序按指数分布递减,则计算公式为:

$$\omega(L_i, C_j) = e^{-\mu(X(L_i, C_j) - 1)} \quad (13)$$

式中:  $X(L_i, C_j)$  为线路  $L_i$  在故障链  $C_j$  中的顺序;  $\mu$  为调节系数,用于调整重要度权重递减的速度,  $\mu$  越小,故障链中的线路权重分布就越均衡。当  $\mu = 0.1$  时,按开断顺序算得的权重依次为 1.000, 0.915, 0.819, 0.741, 0.670 等。

综合来看,线路所涉及的连锁故障越多、引起的故障越严重、开断顺序越靠前,则对应的线路风险重要度也会更高。按  $I(L_i)$  对所有线路进行排序,排名靠前的线路即为连锁故障演化过程中的关键线路。

### 3.2 关键线路辨识流程

关键线路辨识流程如图 4 所示。首先通过改进 DCSS 模型生成故障链集合,然后计算故障链之间的 ED,并通过选取合适的 ED 阈值,应用 AGNES 算法完成对故障链的聚类,分别计算出同类故障链集合中各线路的风险重要度,识别出各类故障演化模式下的关键线路。

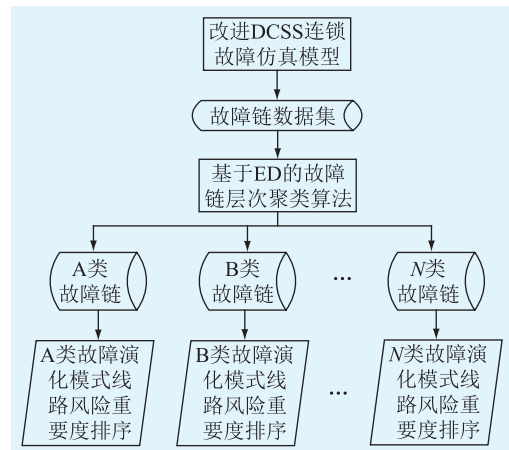


图 4 关键线路辨识流程

Fig.4 Flow chart of critical line identification

## 4 算例分析

为了进一步证明文中算法在大电网中的有效性,文中采用 Matpower 软件包中的算例 case2383wp 进行仿真研究,该算例的详细数据参见文献[23]。

### 4.1 故障链集合生成

文中采用抽样效率较高的 RC 法,通过调用改进 DCSS 连锁故障仿真程序对 Matpower 软件包中的 2 383 节点模型(满足“ $N-1$ ”准则)进行模拟。

2 383 节点模型数据不包含线路平均故障频率等参数,因此文中将 IEEE RTS-79 可靠性标准测试模型的架空线故障率随机映射至文中模型的 2 896 条线路上,使得 2 个模型的线路故障率概率密度分

布函数 $f(P)$ 相近,如图5所示。

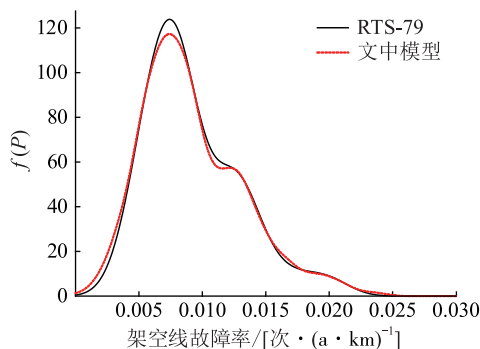


图5 RTS-79与文中模型的线路故障率概率密度曲线对比

Fig.5 Comparison of line failure rate probability density function of RTS-79 and the model in this paper

将映射所得的故障率与实际线路长度相乘得到平均失效频率 $\lambda$ 之后,根据式(5)可算得相应线路的不可用率 $U(L_{init,i})$ ,其中线路平均修复时间 $t_{Dur}$ 取IEEE RTS-79标准数据的中位值10 h。

为平衡故障链搜索时间与搜索广度,文中以负荷损失超过2 701 MW(占总负荷的10%)为筛选阈值,经过10万次RC法循环仿真,共生成424条由不同 $N-2$ 初始故障引发的故障链。对所生成的故障链进行线路频次统计,统计情况如表1所示。

表1 故障链集合中线路出现频次统计

Table 1 Frequency statistics of occurrence of lines in fault chain set

线路出现次数统计/次	线路数量/条	线路出现次数统计/次	线路数量/条
0~25	397	200~225	6
25~50	63	225~250	9
50~75	39	250~275	3
75~100	16	275~300	10
100~125	25	325~350	9
125~150	23	350~375	4
150~175	10	375~400	2
175~200	11		

由表1可知,仅有极少数线路在故障链集合中出现超过300次,而同时有多达397条线路出现小于27次。线路的频次分布具有明显的“长尾效应”,即故障链集合中很少一部分线路的出现次数远高于其他线路,表明这些线路与严重连锁故障的触发与演化存在密切关系。

#### 4.2 故障链聚类结果

对4.1节所得的故障链集合采用ED算法,根据424条故障链之间的相似性可绘制出如图6所示的热力图。以故障链间的ED为依据进行层次聚类,可得如图7所示的聚类过程树状图。

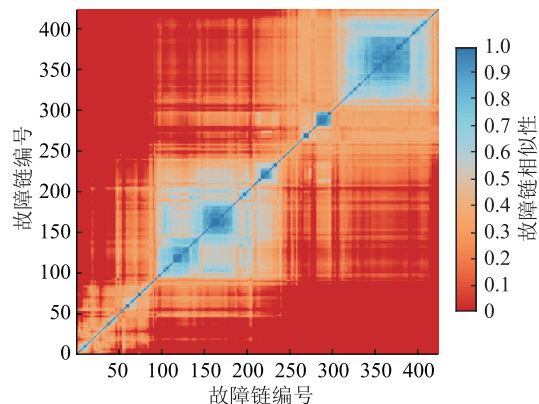


图6 故障链相似性热力图

Fig.6 Fault chain similarity heat map

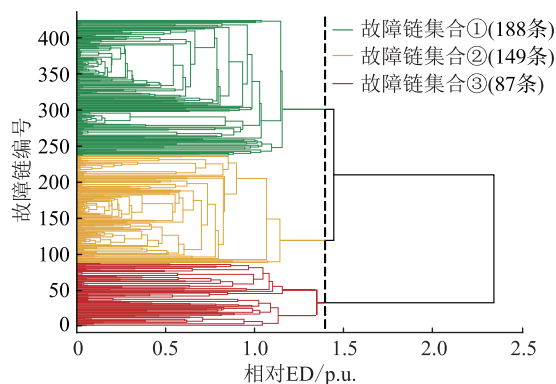


图7 故障链集合层次聚类

Fig.7 Hierarchical clustering of fault chain set

不同簇数的聚类方案所得的CH评估指标如图8所示,由图8可知,当簇数为3时,CH指标最高,表明相应的聚类方案最优。3类故障链已在图7中以不同颜色标注,分别记为故障链集合①~③,同类集合中的故障链相似性较高,表明所涉及的连锁故障演化路径重合度较高。

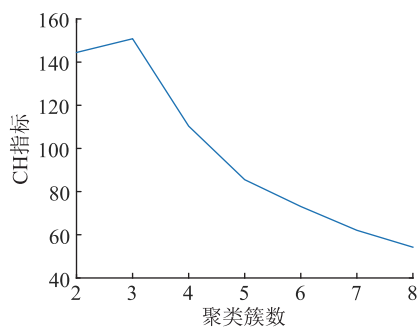


图8 不同聚类簇数对应的CH指标

Fig.8 CH index corresponding to different cluster numbers

#### 4.3 关键线路辨识

对故障链集合①~③分别进行信息统计,所得结果如表2所示,其中平均故障链长度是指集合内每条故障链所开断线路条数的平均值。

表2 不同故障链集合的基本信息统计  
Table 2 Basic information statistics of different fault chain sets

集合	所含故障链数量/条	故障链占比/%	平均故障链长度/条	平均负荷损失/MW
①	188	44.3	102.4	11 299.1
②	149	35.1	60.6	6 274.3
③	87	20.5	37.1	3 461.3

根据式(12)一式(13)分别计算各故障链集合中的线路风险重要度  $I(L_i)$ , 调节系数  $\mu$  取 0.1。将所有线路按  $I(L_i)$  进行降序排列, 排名靠前的线路即为相应故障链集合中的关键线路。表3中分别列出了聚类前后各类故障链集合中线路风险重要度排名前5的关键线路, 括号内的数字表示对应线路在未聚类时故障链总集合中的重要度排名。

表3 不同故障链集合中的关键线路排序  
Table 3 Critical lines rank in different fault chain sets

风险重要度排名	未聚类		聚类后	
	故障链总集合	故障链集合①	故障链集合②	故障链集合③
1	$L_{15}$	$L_{15}(1)$	$L_{250}(12)$	$L_{169}(26)$
2	$L_{96}$	$L_{96}(2)$	$L_{15}(1)$	$L_{52}(11)$
3	$L_{680}$	$L_{113}(4)$	$L_{251}(13)$	$L_{264}(66)$
4	$L_{113}$	$L_{24}(6)$	$L_{771}(7)$	$L_{281}(67)$
5	$L_8$	$L_{680}(3)$	$L_8(5)$	$L_{2528}(70)$

结合表2、表3可知, 由于故障链集合①中的故障链占比较大(44.3%), 若直接对未聚类的故障链总集合进行关键线路辨识, 前5名分别为  $\{L_{15}, L_{96}, L_{680}, L_{113}, L_8\}$ , 此结果与故障集合①中的关键线路前5名  $\{L_{15}, L_{96}, L_{113}, L_{24}, L_{680}\}$  高度重合, 但故障链集合②与③中的关键线路在未聚类故障链总集合中的排名普遍较低, 意味着若不进行聚类极有可能造成故障链集合②与③中关键线路的漏选。

#### 4.4 辨识结果有效性验证

文中参考文献[24], 对算法所辨识的关键线路进行容量扩建, 并基于相同的初始故障集重新进行连锁故障仿真, 扩容后故障链集合的连锁故障风险  $R(S)$  下降得越多, 证明相应算法所辨识的线路在故障链集合中的重要性越高。

线路容量的调整在实际系统中可通过更换线路、增加并行线路等措施实现<sup>[25]</sup>。在文中的算例中, 线路扩建容量设定为额定容量的2倍, 根据4.3节的关键线路辨识结果分别设置以下2组扩容方案。

(1) 扩容方案一。按未聚类时的辨识结果取线路风险重要度排序的前3名, 具体扩容线路为  $\{L_{15},$

$L_{96}, L_{680}\}$ 。

(2) 扩容方案二。按聚类后的辨识结果分别取各故障链集合中的线路风险重要度排序第1名, 具体扩容线路为  $\{L_{15}, L_{250}, L_{169}\}$ 。

另设置随机扩容方案, 即随机选取3条故障链集合中所涉及的线路进行扩容。

对上述3种方案扩容后的2383节点系统重新进行424组连锁故障仿真并进行风险评估, 其中随机扩容的连锁故障风险由20次仿真后取平均值所得, 原系统及3类方案下的风险如图9所示。

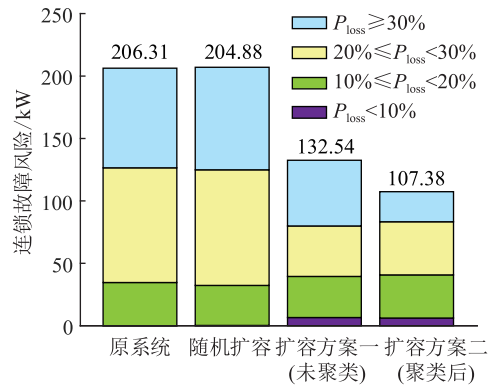


图9 不同扩容方案下系统连锁故障风险比较

Fig.9 Comparison of system cascading failure risks under different expansion plans

由图9可知, 上述的3种扩容方案将连锁故障整体风险由206.31 kW分别降至204.88 kW, 132.54 kW, 107.38 kW。随机扩容方案下, 整体风险平均仅降低了1.43 kW, 证明随机选取的线路重要性较低。这是因为线路的重要性分布与4.1节所述的频次分布类似, 皆具有“长尾效应”, 随机选取线路大概率为重要性不高的非关键线路。而在基于线路风险重要度排序的2种扩容方案中, 方案二能最大程度地降低连锁故障风险, 其中  $P_{\text{loss}} \geq 30\%$  的大规模连锁故障风险由79.8 kW减少至24.1 kW, 明显低于方案一的52.6 kW, 证明分类辨识所得关键线路的重要性更高。

为进一步探究辨识结果对连锁故障风险的影响, 按故障链集合①~③对原系统及扩容方案一、二下的连锁故障风险进行分类统计, 对比结果如表4所示, 其中, 括号内表示相较于原系统连锁故障风险下降的百分比。

由表4可知, 在未聚类时, 方案一使得故障链集合①中的连锁故障风险下降了50.7%, 但同时仅让故障链集合③的连锁故障风险降低了0.36%。因此未聚类时所辨识的线路对大规模故障链集合而言重要性较高, 对其他规模较小故障链集合的影响力则相对有限。相比之下, 方案二下的整体风险下降

表 4 不同扩容方案对特定故障链集合风险的影响  
Table 4 The impact of different expansion schemes on the specific failure chain set risk

故障链集合序号	连锁故障风险/kW		
	原系统	扩容方案一 (未聚类)	扩容方案二 (聚类后)
①	88.69	43.75 (50.7%)	36.73 (58.6%)
②	87.35	53.86 (38.3%)	22.37 (74.4%)
③	30.27	30.16 (0.36%)	26.95 (11.0%)
{①, ②, ③}	206.31	132.54 (35.8%)	107.38 (48.0%)

了48.0%，故障链集合①~③的风险分别下降了58.6%，74.4%，11.0%，皆优于方案一，证明分类辨识的线路集合 $\{L_{15}, L_{250}, L_{169}\}$ 在各类集合中的重要程度皆高于未聚类前的辨识结果。

上述结果说明，若直接对未聚类的424条故障链进行关键线路辨识，辨识结果将倾向于占比较大的故障链集合，难以全面顾及各类连锁故障演化模式，所辨识的线路在规模较小集合中的重要性也相对较低。相比之下，基于故障链聚类算法的分类辨识结果使得各类连锁故障风险的下降程度更高，进一步证明了故障链聚类算法对于提高关键线路辨识效果的有效性。

#### 4.5 与其他关键线路辨识算法的比较

为进一步验证文中所提关键线路分类辨识算法的有效性，文中针对2383节点模型采用潮流系数法<sup>[26-27]</sup>与连锁故障关系图(cascading failure graph, CFG)法<sup>[25]</sup>，分别得到线路重要性排序，根据排序结果前3名分别设置相应的线路扩容方案。其中潮流系数法具体扩容线路为 $\{L_{169}, L_{52}, L_{23}\}$ ，CFG法具体扩容线路为 $\{L_8, L_{38}, L_{90}\}$ 。对扩容后的系统重新进行仿真与连锁故障风险评估，所得结果与4.4节的扩容方案一、二比较，对比结果如图10所示。

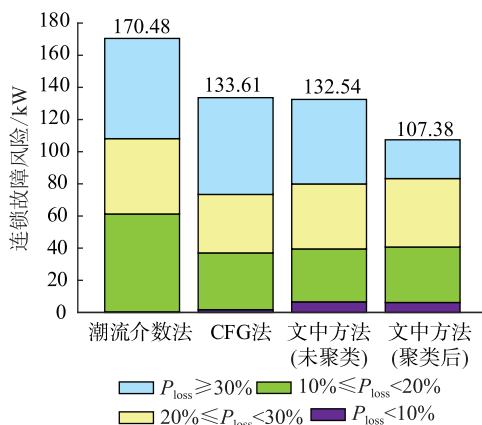


图 10 不同算法所辨识线路扩容后的连锁故障风险比较  
Fig.10 Comparison of cascading failure risks of lines identified by different algorithms after capacity expansion

由图10可知，由于潮流系数法未考虑连锁故障的演化过程，因此辨识结果对降低系统连锁故障风险的作用较小；CFG法与文中所提方法(未聚类)虽能一定程度降低系统连锁故障风险，但缺乏对各类连锁故障演化模式的综合考虑，作用依然有限；相比之下，基于故障链聚类的关键线路扩容方案能够最大程度地降低系统的连锁故障风险，证明文中方法(聚类后)所辨识线路的重要程度更高。

## 5 结语

文中首先建立了改进DCSS仿真模型，并对过载主导型连锁故障进行仿真，生成含丰富时序信息的故障链集合。针对现有关键线路辨识算法未能对各类故障演化模式予以全面考虑的不足，文中提出了基于ED的故障链聚类算法，对相似度较高的故障链集合进行分类评估，能够更加精确地揭露特定演化路径下的薄弱环节，可进一步提高关键线路的辨识精度与效率。以Matpower 2383节点系统为例，通过对关键线路进行扩建，以扩容前后的连锁故障风险水平为依据，量化比较了各类算法所辨识线路的重要性，进一步证明了所提模型及算法的有效性。文中的研究工作还可从以下方面进一步完善和深入：

- (1) 随着高压直流的投运及大量电力电子器件在电力系统中应用，连锁故障过程中的无功问题、交直流耦合问题日益突出，基于更复杂模型及其故障链数据集的关键线路辨识将是后续研究的要点。
- (2) 文中从故障演化路径相似性的角度进行聚类，后续可进一步考虑如失负荷量、系统解列程度等其他变量作为聚类依据。

#### 参考文献:

[1] 李保杰,李进波,李洪杰,等.土耳其“3.31”大停电事故的分析及对我国电网安全运行的启示[J].中国电机工程学报,2016,36(21):5788-5795,6021.  
LI Baojie, LI Jinbo, LI Hongjie, et al. Analysis of Turkish black-out on March 31, 2015 and lessons on China power grid [J]. Proceedings of the CSEE, 2016, 36(21): 5788-5795, 6021.

[2] 汤涌,卜广全,易俊.印度“7.30”、“7.31”大停电事故分析及启示[J].中国电机工程学报,2012,32(25):167-174,23.  
TANG Yong, BU Guangquan, YI Jun. Analysis and lessons of the blackout in Indian power grid on July 30 and 31, 2012 [J]. Proceedings of the Chinese Society for Electrical Engineering, 2012, 32(25): 167-174, 23.

[3] 严道波,文劲宇,杜治,等.2021年德州大停电事故分析及其对电网规划管理的启示[J].电力系统保护与控制,2021,49(9):121-128.  
YAN Daobo, WEN Jinyu, DU Zhi, et al. Analysis of Texas

- blackout in 2021 and its enlightenment to power system planning management[J]. *Power System Protection and Control*, 2021, 49(9):121-128.
- [4] YANG Y F, GUAN X H, ZHAI Q Z. Fast grid security assessment with  $N-K$  contingencies[J]. 2017 IEEE Power & Energy Society General Meeting, 2017:1.
- [5] 梁颖, 方瑞明, 卢小芬, 等. 基于复杂网络理论的电力系统脆弱线路辨识研究现状[J]. *电力系统保护与控制*, 2012, 40(20):151-155.  
LIANG Ying, FANG Ruiming, LU Xiaofen, et al. Research status of power system vulnerable line identification based on complex-network theory[J]. *Power System Protection and Control*, 2012, 40(20):151-155.
- [6] 吴晨, 韩海腾, 祁万春, 等. 基于改进线路效能的电网脆弱性辨识方法研究[J]. *电力工程技术*, 2018, 37(5):64-68.  
WU Chen, HAN Haiteng, QI Wanchun, et al. Vulnerability identification in power systems based on an improved line-efficiency evaluation approach[J]. *Electric Power Engineering Technology*, 2018, 37(5):64-68.
- [7] 李勇, 刘俊勇, 刘晓宇, 等. 基于潮流熵的电网连锁故障传播元件的脆弱性评估[J]. *电力系统自动化*, 2012, 36(19):11-16.  
LI Yong, LIU Junyong, LIU Xiaoyu, et al. Vulnerability assessment in power grid cascading failures based on entropy of power flow[J]. *Automation of Electric Power Systems*, 2012, 36(19):11-16.
- [8] 张振安, 郭金鹏, 张雪敏, 等. 考虑频率稳定的大停电事故模型及应用[J]. *电力系统及其自动化学报*, 2015, 27(4):26-32.  
ZHANG Zhen'an, GUO Jinpeng, ZHANG Xuemin, et al. Black-out model and its application considering frequency stability[J]. *Proceedings of the CSU-EPSA*, 2015, 27(4):26-32.
- [9] ALIZADEH MOUSAVI O, CHERKAOUI R, BOZORG M. Black-outs risk evaluation by Monte Carlo simulation regarding cascading outages and system frequency deviation[J]. *Electric Power Systems Research*, 2012, 89:157-164.
- [10] EPPSTEIN M, HINES P. A "random chemistry" algorithm for identifying collections of multiple contingencies that initiate cascading failure[J]. 2013 IEEE Power & Energy Society General Meeting, 2013:1.
- [11] 肖遥, 过羿, 谢佳, 等. 基于 Apriori 算法的连锁故障机理研究[J]. *电力学报*, 2017, 32(1):22-27, 69.  
XIAO Yao, GUO Yi, XIE Jia, et al. Research on the mechanism of cascading failure based on Apriori algorithm[J]. *Journal of Electric Power*, 2017, 32(1):22-27, 69.
- [12] 周强明, 姜盛波, 鲁鸿毅, 等. 基于 FP-growth 算法的大电网关键线路辨识方法[J]. *电力自动化设备*, 2018, 38(4):89-95.  
ZHOU Qiangming, JIANG Shengbo, LU Hongyi, et al. Critical line identification method based on FP-growth algorithm for large power grid[J]. *Electric Power Automation Equipment*, 2018, 38(4):89-95.
- [13] 刘豫航, 黄少伟, 梅生伟, 等. 基于序列模式挖掘的电力系统连锁故障模式分析方法[J]. *电力系统自动化*, 2019, 43(6):34-40.  
LIU Yuhang, HUANG Shaowei, MEI Shengwei, et al. Analysis on pattern of power system cascading failure based on sequential pattern mining[J]. *Automation of Electric Power Systems*, 2019, 43(6):34-40.
- [14] 陈海燕, 刘晨晖, 孙博. 时间序列数据挖掘的相似性度量综述[J]. *控制与决策*, 2017, 32(1):1-11.  
CHEN Haiyan, LIU Chenhui, SUN Bo. Survey on similarity measurement of time series data mining[J]. *Control and Decision*, 2017, 32(1):1-11.
- [15] ANGHEL M, WERLEY K A, MOTTER A E. Stochastic model for power grid dynamics[C]//2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07). Waikoloa, HI, USA. IEEE, 2007:113.
- [16] REZAEI P, HINES P D H, EPPSTEIN M J. Estimating cascading failure risk with random chemistry[J]. *IEEE Transactions on Power Systems*, 2015, 30(5):2726-2735.
- [17] 李文沅. 电力系统风险评估:模型方法和应用[M]. 北京:科学出版社, 2006.  
LI Wenxuan. Power system risk assessment: model method and application[M]. Beijing: Science Press, 2006.
- [18] 邱海峰, 陈兵, 袁晓冬, 等. 基于动态时间弯曲距离的电压暂降源辨识方法[J]. *电力系统保护与控制*, 2017, 45(13):7-13.  
QIU Haifeng, CHEN Bing, YUAN Xiaodong, et al. Identification of voltage sag sources based on DTW[J]. *Power System Protection and Control*, 2017, 45(13):7-13.
- [19] 吴杰, 吕春媚, 崔彪. 基于小波轮廓描述符和动态时间扭曲的电压扰动分类[J]. *电工电能新技术*, 2015, 34(4):62-67, 80.  
WU Jie, LYU Chunmei, CUI Biao. Identification of disturbance signals based on wavelet boundary descriptors and dynamic time warping[J]. *Advanced Technology of Electrical Engineering and Energy*, 2015, 34(4):62-67, 80.
- [20] RISTAD E S, YIANILOS P N. Learning string-edit distance[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(5):522-532.
- [21] 孙即祥. 现代模式识别[M]. 2版. 北京:高等教育出版社, 2008.  
SUN Jixiang. Modern pattern recognition[M]. 2nd ed. Beijing: Higher Education Press, 2008.
- [22] CALINSKI T, HARABASZ J. A dendrite method for cluster analysis[J]. *Communications in Statistics-Simulation and Computation*, 1974, 3(1):1-27.
- [23] RAY Z. CASE2383WP power flow data for polish system-winter 1999-2000 peak[EB/OL]. (2018-10-09) [2021-04-10]. <https://github.com/MATPOWER/matpower/blob/master/data/case2383wp.m>.
- [24] LI L Z, WU H, SONG Y H, et al. A state-failure-network method to identify critical components in power systems[J]. *Elec-*



tric Power Systems Research, 2020, 181:106192.

[25] HINES P D H, DOBSON I, REZAEI P. Cascading power outages propagate locally in an influence graph that is not the actual grid topology[J]. IEEE Transactions on Power Systems, 2017, 32(2):958-967.

[26] 刘文颖, 梁才, 徐鹏, 等. 基于潮流介数的电力系统关键线路辨识[J]. 中国电机工程学报, 2013, 33(31):90-98, 11. LIU Wenying, LIANG Cai, XU Peng, et al. Identification of critical line in power systems based on flow betweenness[J]. Proceedings of the CSEE, 2013, 33(31):90-98, 11.

[27] 陶鸿飞, 谢栋, 赵福林, 等. 含大规模风电并网电力系统的关键线路综合辨识[J]. 电力系统保护与控制, 2020, 48(6):115-123.

TAO Hongfei, XIE Dong, ZHAO Fulin, et al. Comprehensive identification of critical line in power systems with large-scale wind power integration[J]. Power System Protection and Control, 2020, 48(6):115-123.

作者简介:



黎寿涛

黎寿涛(1996),男,硕士在读,研究方向为连锁故障仿真及风险评估(E-mail:ep\_lishoutao@mail.scut.edu.cn);

夏成军(1974),男,博士,副教授,研究方向为电力系统稳定分析与控制;

钟明明(1996),男,硕士在读,研究方向为电力系统分析。

### Critical line identification of power grid based on fault chain clustering algorithm

LI Shoutao<sup>1,2</sup>, XIA Chengjun<sup>1,2</sup>, ZHONG Mingming<sup>1,2</sup>, GUAN Lin<sup>1,2</sup>

(1. School of Electric Power, South China University of Technology, Guangzhou 510640, China;

2. Guangdong Province Key Laboratory of Intelligent Operation and Control for New Energy Power System, Guangzhou 510663, China)

**Abstract:** Research on the development mechanism of cascading failure and identification the critical line of evolution path is of great significance for exposing weak links in the power system and reducing the risk of cascading failures. For this reason, an identification method for critical lines of power grid based on fault chain clustering algorithm is proposed. Firstly, an improved direct current power flow simulator of power system separation (DCSS) cascading failure simulation model, which combines with the random chemistry (RC) method to efficiently generate fault chain set with detailed timing information is established. Then, the hierarchical clustering of the failure chain set is realized by using edit distance as the similarity index, and the classified fault chain set can not only reduce the difficulty of subsequent data mining, but also more comprehensively identify the weak links in different cascading failure evolution modes. Finally, taking the Matpower 2 383 node system as an example, the importance of the lines identified by various algorithms is evaluated through the reduction of the system risk level after the expansion of critical lines. The results show that the proposed method can better reduce the risk level of cascading failure, which proves the effectiveness of the proposed model and algorithm.

**Keywords:** cascading failure; similarity measurement; edit distance; hierarchical clustering; risk assessment; critical line identification

(编辑 陆海霞)